The Puzzle of Missing Female Engineers: Academic Preparation, Ability Beliefs, and Preferences*

Ying Shi†

Stanford University

Abstract

This paper uses administrative North Carolina data linked from high school to college and national surveys to characterize the largest contributor to the STEM gender gap: engineering. Disparities are the result of differential entry during high school or earlier rather than postsecondary exit. Differences in pre-college academic preparation account for 5 to 7% of the gap. Females' relative lack of academic self-confidence explains 8%, while other-regarding preferences and professional goals capture a further 14%. Empirical evidence using identifying variation in the gender composition of twins in North Carolina shows that opposite-sex pairs are more likely to pursue gender-stereotypical majors.

Keywords: Economics of gender, human capital, occupational choice

JEL Classification Numbers: J16, J24, I23

*Citation: Shi, Y. (2018) "The Puzzle of Missing Female Engineers: Academic Preparation, Ability Beliefs, and Preferences" *Economics of Education Review* 64: 129-143.

†Center for Education Policy Analysis; 520 Galvez Mall, CERAS Building, Stanford University, Palo Alto, CA 94305 (E-mail: yingsh@stanford.edu; Phone: 281-687-0926). The author is grateful to Peter Arcidiacono, Charles Clotfelter, V. Joseph Hotz, Helen Ladd, Seth Sanders, Jacob Vigdor, and seminar participants at the AEFP annual conference for their helpful comments. This work has been supported by an Institute of Education Sciences Postdoctoral Training Fellowship Award (R305B130017), the American Education Research Association (AERA) Minority Dissertation Fellowship and a Horowitz Foundation for Social Policy grant.

1 Introduction

The past several decades bore witness to significant changes in education-related gender gaps. In the years immediately following World War II, only one female enrolled in college for every 2.3 males. Yet by the 1980s, women surpassed men in both college enrollment and completion (Goldin et al., 2006). These successes in postsecondary educational attainment, however, failed to translate into higher rates of female participation in select science, technology, engineering and math (STEM) fields, where women are still vastly underrepresented (Turner and Bowen, 1999; Griffith, 2010). In particular, gender disparities are most glaring in the subfield of engineering, where women comprise only 12 percent of working engineers in 2013 (Legewie and DiPrete, 2014; Corbett, 2015). The persistence of the sizable engineering disparity in spite of gender gap reversals in non-STEM subjects and STEM ones such as biology is a prevailing puzzle.

Gender disparities in fields of study have lasting consequences for longer-term earnings and skill distributions. STEM graduates enjoy a substantial pay premium relative to peers in other fields. The difference in log wages between engineers and education majors, for instance, rivals the earnings gap between college and high school graduates (Altonji et al., 2012). Differential take-up of science and math-intensive fields account for a notable share of the male-female earnings gap, such that achieving gender parity on major choice could significantly reduce earnings inequality (Paglin and Rufolo, 1990; Brown and Corcoran, 1997; Blau and Kahn, 2000). Gender gaps in major or occupational choice can also lead to differential accumulation of STEM-focused human capital among men and women that matter for tomorrow's workforce.

This paper uses new administrative data from North Carolina and a pooled national survey of college freshmen to investigate the largest contributor to STEM disparities: engineering. While a plethora of economics studies focuses on the aggregate STEM gender gap, comparatively little research examines specific STEM subfields. Yet the divergent patterns by subfield, from postsecondary gender parity in biology to striking gaps in computer science and engineering, necessitate a more targeted approach. This work contributes evidence on the gender gap along three dimensions. The first is to document the size of the disparity in engineering and its evolution from the beginning of high school through postsecondary schooling. Linked administrative data permits a detailed look at how major orientation in high school translates to actual major choice during the critical transition to college. Second, I differentiate between the roles of entry versus exit during college, using administrative postsecondary data to document attrition rates by gender. The final and most substantial component examines contributors to the gender gap, ranging from individuals' ability beliefs to professional preferences. The combination of a statewide longitudinal dataset and national survey data enables a more comprehensive account of factors underlying the engineering gender gap than previously available.

The datasets' temporal coverage permits a closer look at engineering participation starting in high school. Using engineering orientation or choice as outcome variables, I document a disparity of over 8 percentage points in 9th grade and 11 percentage points after the first year of postsecondary education. The magnitude of this gap is especially striking in light of baseline female engineering participation rates between 2 to 4%. Longitudinal data in North Carolina shows that while the majority of the postsecondary gender gap is explained by high school engineering orientation, women are nevertheless less likely to convert early interest in engineering to actual major choice. Once students have declared an engineering major in the North Carolina public university system, I find no evidence to support systematically higher attrition among female students. Results indicate that the gap is mainly attributable to lower entry among female students rather than higher exit during this period. Efforts to increase the rate of female entry and reduce gender divergence in STEM orientation, in particular engineering, should begin no later than high school and not neglect the crucial transition into college.

Tailored policies rely on a better understanding of the gender gap's contributors. I investigate four explanatory accounts: differences in academic preparation, differences in academic ability beliefs, differences in prosocial values and professional goals, and the role of family structure and gender-based norms. Decomposition evidence shows that SAT scores and high school GPA account for between 5 to 7% of the overall disparity. Course-taking patterns in the first half of high school betray few clues on eventual major orientation. Meanwhile, beliefs about lower academic ability dissuade women from entering the field even after controlling for academic performance. Elevating women to the same belief levels as their male counterparts would bridge the gender gap by 8%. Female preferences for prosocial responsibilities and contributing to the arts over sciences explain over 14% of the gap. Notably, decomposition results for the full sample disguise substantial heterogeneity across racial groups and baseline math ability. Explanatory factors collectively explain more of the gender gap for white, Hispanic, and Asian students than African American students because black females track their male peers more closely in academic preparation and professional goals. The importance of ability beliefs and professional goals is also increasing in baseline math ability. Overcoming the female math confidence deficit alone would bridge the gender gap by 7% among students most academically prepared to enter engineering, relative to 4-5% among lower-scoring students.

Complementing these explanatory accounts is a set of gender-specific norms and expectations conveyed mainly in the family context by parents. To better understand their influence on STEM participation, I use a sample of twin pairs in North Carolina under the identifying assumption that sex assignment is exogenous. A potential challenge to this assumption is the inability to distinguish identical twins from fraternal twins in the data. As such interpretation depends on how unobserved genetic or environmental factors affecting the presence of identical twins might be correlated with future parental expectations and investments. I find that males from opposite-sex pairs are substantially more likely than males from same-sex pairs to choose engineering as a preferred major. These results cannot be explained by differential math ability or a relative advantage story, in which STEM is chosen by the twin with higher math performance. The specialization along gender-stereotypical lines suggests that gender roles and expectations can play a meaningful role, for instance by encouraging boys to invest in more male-dominated pursuits such as computer skills and engineering.

This paper is organized as follows. The next section highlights engineering's contribution to the overall STEM gap and grounds this study in related literature. Section 3 details the three main administrative and cross-sectional datasets utilized for decomposition. Section 4 describes the role of entry vs. exit, while the subsequent section outlines the empirical strategy. Section 6 presents evidence on the relevance of each explanatory account. I conclude with a discussion of implications.

2 Factors contributing to the STEM gender gap

Gender gaps in major choice are large and persistent in the US context. Table 1 uses Census data to document the share of recent college graduates across all STEM subjects and by subfield.[1] Although males are twice as likely to graduate from college with a STEM degree on aggregate, this result disguises large variations by subfield. Degree attainment in biology nears gender parity, while fields such as engineering and computer science still exhibit sizable gaps. 11.7% of male graduates select engineering, compared to only 2.5% of females.[2] Over 9 percentage points of the 16 percentage point STEM gap is attributable to gender disparities in engineering, while computer science contributes an additional 4

[1] I define a field as STEM if it belongs to one of the following categories: 1) Agriculture, 2) Computer and Information Sciences, 3) Engineering, 4) Engineering Technologies, 5) Biology and Life Sciences, 6) Mathematics and Statistics, 7) Physical Sciences, and 8) Nuclear, Industrial Radiology, and Biological Technologies, abbreviated as Science Tech.

[2] While this study focuses on the US context, its findings are consistent with gender disparities documented in other countries. For instance, Card and Payne (2017) find a 13.2 percentage point gender difference in engineering participation among Canadian workers between the ages of 25 and 34.

percentage points. Since engineering plays an outsized role in informing the STEM gender gap, it is the central focus of this paper.

⟨ Table 1 about here ⟩

In decomposing the gender gap into explanatory accounts, I draw upon a wealth of literature exploring cross-gender differences in the STEM context. Previous research concentrates on several sources of disparity: academic skills and preparation, family background and expectations, tastes or preferences such as those related to pecuniary payoffs or the work environment, and psychosocial attributes such as ability beliefs. Within academic preparation, the preponderance of research focuses on math skills. Earlier studies often cite differences, although gaps in performance have closed in recent years (Xie and Shauman, 2003; Hyde et al., 2008). Recent studies find small or insignificant gender differences in math standardized tests across elementary and secondary schools (Hyde et al., 2008; Hyde and Mertz, 2009; Sass, 2015), while others find that gaps only materialize several years into school entry (Fryer and Levitt, 2010). Among the mathematically gifted, evidence for higher variability among males shows diminishing gaps over time.[3] Conditioning on performance and grades still leaves a large unexplained residual in the STEM gender gap, suggesting that academic preparation plays a relatively minor role (Turner and Bowen, 1999; DiPrete and Buchmann, 2013; Card and Payne, 2017).

Family background is another potential source of influence on individuals' STEM orientation. Parental expectations of children's math and science abilities and academic trajectories may differ by child's gender, thereby affecting students' investments in such skills (Eccles et al., 1990). These expectations can be shaped by parents' own educational and occupational experiences. While a growing body of literature is formally incorporating parental beliefs as an input into human capital production, limited empirical evidence exists on the role of parental influence for STEM orientation (Fryer and Levitt, 2010; Agostinelli and Wiswall, 2016).

Pre-labor market skill accumulation also depends on individual preferences for field or job attributes. Women may enjoy taking non-STEM courses and sort into those fields on the basis of non-pecuniary factors. Over time, differences across gender preferences can lead to clearly differentiated human capital acquisition. There is growing evidence affirming the important role of preferences. Zafar (2013), for instance, finds that differences in coursework and workplace enjoyment and gaining parents' approval are the primary

[3] A 13:1 ratio of men to women among high SAT math achievers in the early 1980s has since bridged to approximately 2:1 at the top end of the distribution (Benbow and Stanley, 1983; Ellison and Swanson, 2010).

explanations of divergent major choices among male and female college students.[4]

One explanatory account receiving increasing attention is psychosocial attributes. There is accumulating evidence that gender-based differences in confidence and mindsets influence individuals' academic behavior (Sax, 1994; Beyer and Bowden, 1997; Dar-Nimrod and Heine, 2006). Laboratory experiments in economics show that conditional on performance, higher confidence can drive men to enter competitive arenas at greater rates than women (Gneezy et al., 2003; Niederle and Vesterlund, 2007). In the field, psychosocial attributes can affect actual academic decisions. Confidence and an inclination for competitiveness influence females' performance in high-level math tests (Niederle and Vesterlund, 2010). Females who ascribe to a fixed view of ability do worse than those who emphasize an experiential and malleable account of math ability (Dar-Nimrod and Heine, 2006).[5] The preponderance of math-based curricula in STEM majors brings these issues to the fore. Insofar as students perceive STEM majors to require technical mastery, gender gaps in beliefs about one's own ability can lead men and women to sort into different academic tracks (Correll, 2004).

While the above accounts have been tested in a wide number of studies to illuminate STEM gender disparities, significantly less is known about explanatory accounts for any given subfield such as engineering. Existing studies predominantly focus on the postsecondary engineering gap and document the roles of factors such as achievement beliefs or academic performance using survey data (Vogt et al., 2007; Heyman, 2002; Sax et al., 2016). This paper uses statewide administrative data and two national surveys to comprehensively document the evolution of the engineering gender gap from secondary schooling through college. Data from high school years informs the appropriate timing of interventions, as earlier policy responses are necessary if STEM orientation is diverging in lower grades. Another advantage of the present study is the availability of longitudinal data on

[4] An important question beyond the scope of this paper is how these preferences develop and evolve over the life course. Evidence shows that environmental factors such as academic context matter. Attending single-sex schools or classrooms with higher shares of females can encourage more women to choose STEM majors (Solnick, 1995; Billger, 2002; Favara, 2012). Similarly, exposure to female teachers and faculty can increase female students' participation in STEM courses and majors (Rothstein, 1995; Bettinger and Long, 2005; Dee, 2007; Carrell et al., 2010), although some studies find non-existent or only temporary effects (Canes and Rosen, 1995; Sass, 2015). Thus the social or institutional context can shape gender gaps by influencing individual preferences and subsequently affecting students' investments in science and math skills. In this respect, preferences are not independently determined, but rather dynamically connected to academic preparation and social context in its development.

[5] Research on mindsets finds that women are more likely to hold a fixed view of intelligence, where ability is intrinsic and cannot be easily gained, while males are more likely to ascribe to an incremental theory of intelligence that enables augmentation through hard work (Dweck, 2000, 2008). The specific academic context can interact with and activate these attitudes. A recent paper found that academic fields that believe intrinsic, raw talent are important for success exhibit particularly large gender disparities (Leslie et al., 2015).

major orientation from the end of high school to college. This permits a study of changes in STEM orientation and factors underlying switches during an important transition period.

3 Data and descriptive statistics

3.1 North Carolina high schools

The first dataset comprises administrative records spanning all public and charter secondary schools from the North Carolina Education Research Data Center (NCERDC). Beginning in 2009, NCERDC supplemented the database with College Board data on SAT scores and major intentions at the end of high school. One set of outcome variable derives from students' responses to a question on "First Choice Major." Answers on preferred major and the certainty of this choice are recorded during the latest administration of the SAT taken by the high school student.[6] I limit the sample to students who do so during their junior or senior year, such that the variable describes major intentions during the second half of high school.[7] Unique data on actual college behavior is available in 2010, when the College Board begins listing all colleges to which the student submitted their SAT score reports.

Students' academic achievement variables derive from high school transcript files that detail courses taken and grades associated with each. It is possible to construct cumulative GPA for the first two years of high school using course-level data. I also include earned credit hours in reading, math, physical science, and computer programming during the first half of high school.[8] The longitudinal combination of students' academic achievement history and forward-looking plans render these data elements suitable for exploring academic factors associated with major orientation. In addition, the SAT questionnaire solicits information on extracurricular activities in grades 9 and 10 including participation in computer and musical activities.

Table A1 summarizes SAT and GPA performance, earned course credits, and extracurricular participation using the 2009 - 2014 cohorts of graduating seniors. It differentiates

[6] Options given in the SAT questionnaire on the certainty of the student's first choice major include "very certain," "fairly certain," and "not certain."

[7] One caveat is that the outcome is conditional on taking the SAT. Students who do not intend to enroll in a 4-year institution or took the ACT in place of the SAT are excluded from this sample. Using the 2009 cohort, 47% of high school seniors in the NCERDC database took the SAT at least once.

[8] Each student is given a 2-year period for accumulating credits in each subject under the assumption of regular academic promotions. For example, course history information for a graduating senior in 2009 derives from 2007 10th grade and 2006 9th grade transcript files. The physical science category describes cumulative credits earned in physical science, chemistry, and physics courses.

the academic trajectories of those not inclined towards engineering from same-gender peers who prefer this track. For example, aspiring engineers of both genders differentiate themselves by earning more physical science and computer programming credits and fewer reading credits in the first half of high school. While non-engineers earn more math credits early on, the credit advantage reverts back to aspiring engineers by 10th grade. On the extracurricular front, participation in computer activities are noticeably higher among the engineering-oriented. Aspiring female engineers are overall more selected in academic ability than male counterparts. While female non-engineers have lower SAT math and verbal scores than male non-engineers, females who indicate an engineering interest have the highest SAT math and verbal scores of any group.

In addition to detailed academic data, NCERDC files provide some context on family structure by flagging twin pairs using identifying information such as name and birth date. I construct a sample of twins that ever enrolled in a North Carolina public or charter school between 2004 and 2007. Of the 9570 pairs in the full sample, 35% are opposite-sex twins, with the remaining split between same-sex female twins and same-sex male twins (Table A4). The analytic sample includes pairs who are matched to 2009 - 2014 College Board data with non-missing SAT scores.

3.2 University of North Carolina

A particular advantage of North Carolina data is the ability to track students from public school entry through college graduation in the state's public university system. An unique identifier links primary and secondary school records to administrative data in the 16-campus University of North Carolina (UNC) system. I focus on the 2010 cohort due to coverage by both College Board and UNC data. The availability of major-related variables renders this longitudinal dataset well suited for studying the relationship between major orientation and actual choice as students navigate the transition to postsecondary.

The base sample includes students in a North Carolina public high school during 2010 who took the SAT exam before enrolling in a UNC institution. I exclude unmatched observations that correspond to out-of-state students, in-state residents who attended private school, and public school attendees with missing SAT information. Major orientation comes from responses to the SAT questionnaire, while postsecondary major choice derives from enrollment records covering students' term-by-term credit accumulation, GPA and declared major. Using enrollment data, I can track students' declared majors at a point in time or for a given course credit milestone.

Students expressing an engineering orientation during high school but end up choosing a different major in college can depart from their original intention at several points. Data on college application portfolios provides information on whether application rates to UNC campuses with engineering undergraduate degrees differs by gender. UNC transcript data shows final UNC enrollment and course sequences taken by term. To better understand the extent to which major orientation translates to actual major choice and how this can differ along gender lines, I follow a sample of engineering-oriented students through UNC enrollment. Summary statistics in A2 juxtapose the academic achievements, college application and UNC curricular exposure of the high school students that chose engineering as their preferred major. Female students in this sample have similar SAT math scores as their male peers, but exhibit higher SAT verbal scores and cumulative GPA.

3.3 CIRP Freshmen Survey

The third dataset is an annual survey of entering full-time college freshmen administered by the Cooperative Institutional Research Program (CIRP). As the largest continuous national survey of college students, the CIRP Freshmen Survey provides a snapshot of incoming students' background characteristics and college expectations.[9] The dependent variable on engineering intentions comes from a question eliciting students' probable fields of study. Family background variables include parents' occupational categories and total income, while measures of academic preparation derive from student self-reports of SAT and ACT scores and high school GPA.[10]

A key advantage of the survey is the breadth of its coverage spanning multiple factors of potential relevance to the gender gap. The first involves self-confidence and academic ability beliefs in the form of self-reported assessments of academic, mathematical, and writing abilities.[11] Conditioning on standardized test scores, high school GPA, and other objective measures of academic performance allows for examining the role of academic self-confidence independent of assessed academic ability. Apart from ability beliefs, some students may prefer engineering for its expected pecuniary benefits or compatibility with preferences for problem-solving and scientific inquiry. These tastes and preferences are partially captured via questions on personal goals and expected future acts.[12] The survey

[9]Most universities and colleges administer the survey during student orientation, although the survey is typically made available between March and October annually.

[10]Previous literature on the relationship between self-reported and actual GPAs finds reasonable validity, with a particularly strong positive correlation for higher ability students (Kuncel et al., 2005).

[11]Students are asked to rate themselves on each trait along a scale of 1) lowest 10%, 2) below average, 3) average, 4) above average, and 5) highest 10%.

[12]Students evaluate the personal importance of each social, political, academic or economic goal by de-

incorporates economic considerations via a question on the importance of "being very well off financially." Coverage of the role of prosocial and other-regarding values includes questions on the importance of "helping others who are in difficulty," "influencing social values," and likelihood that the student will "participate in volunteer or community service work." Professional goals in the arts and sciences are captured by the importance of "creating artistic works" and "making a theoretical contribution to science." Finally, the relationship between family considerations and major choice relies on a variable on the importance of raising a family.

While the CIRP Freshmen Survey extends as far back as 1965, I constrain the analytic sample to more recent cohorts to ensure the consistency of survey variables over time. The sample retains all students in four-year colleges or universities with non-missing demographics information, SAT scores, and parental occupational categories.[13] The pooled cross-sectional base sample of 2,042,832 students spans the 1990-1999, 2001, 2004, 2006, 2008 and 2010 cohorts. Average SAT scores in the survey are higher than the UNC sample, reflecting likely compositional differences in the sample of participating universities and students (Table A3). Aspiring engineers are highly selected on attributes such as mathematical ability beliefs and interest in making a theoretical contribution to science.

4 The engineering pipeline: entry vs. exit

Table 2 traces the engineering gender gap from the beginning of high school to post-secondary education. Among multiple data sources documenting sizable and persistent gaps in engineering orientation, the earliest data point comes from the High School Longitudinal Study of 2009 (HSLS:09). The HSLS:09 uniquely provides information on early engineering predilection for a nationally representative sample of 9th graders that is absent in administrative data and other recent longitudinal surveys. Results show that 9th grade students exhibit a 8.2 percentage point gender disparity in engineering orientation, defined by a preference for an engineering job or occupation at age 30.[14] Students in the

scribing them as 1) not important, 2) somewhat important, 3) very important, and 4) essential. The survey also asks students to guess the probability of undertaking a future action, such as changing their major or dropping out of college. Students choose between 1) no chance, 2) very little chance, 3) some chance, and 4) very good chance.

[13] For the small subset of students with only ACT composite scores and missing SAT values, I input SAT math and verbal scores using the sample of individuals who took both standardized exams and a quartic of ACT scores. Across the sample, less than 5% of academic beliefs, personal goals, and expected future acts covariates have missing values. I include indicators for missing data and sample means in place of missing data.

[14] The outcome variable is constructed from student responses to the question "what is the job or occupation that you expect or plan to have at age 30?" during the fall of their high school freshman year. Students

second half of high school exhibit a 14.7 percentage point gap, comparable to the 14.0 percentage gap documented at the beginning of students' postgraduate careers using the CIRP Freshmen Survey. One concern with high school data is that major intentions indicated on a college entrance exam may not endure into college and translate into actual major choice. The last specification follows students' decisions using UNC administrative files that document term-by-term major choice. Column (4) presents a snapshot of the full sample of UNC students who attained at least 30 credit hours (approximately one year of study) at their home institutions. Among this group, females are 11.5 percentage points less likely to choose engineering. The magnitude of this disparity becomes more apparent when it is compared to baseline participation rates: just over 3 percent for women.

⟨ Table 2 about here ⟩

These sizable gender gaps in STEM participation are consistent with several explanations. First, they could be attributable to differences in the sampled population, from the nationally representative HSLS:09 to aspiring college students in North Carolina. To gauge the extent of sample selection in the North Carolina context, I juxtapose the full sample of high school seniors with SAT test-takers aspiring to college and students who eventually enroll in the public university system. Table B1 shows that females are overrepresented among SAT test-takers relative to the full high school sample at 55% and 51%, respectively.[15] Conditional on taking the test, the share of females enrolling in UNC is slightly larger relative to female representation among SAT test-takers (Table B2).[16] One question that arises is whether differential selection into college explains a portion of the STEM gender gap (Card and Payne, 2017). This is ostensibly not the case for engineering, since female enrollment in the five main UNC engineering campuses is close to parity.[17]

Another potential difference is in the way each sample solicited engineering orientation. Question wording in the HSLS:09 referred to expected job or occupation, while

are categorized as having an engineering orientation if they choose the "Architecture and Engineering" occupation category with the STEM sub-domain exclusively in Life and Physical Science, Engineering, Mathematics, and Information Technology Occupations. All students with STEM sub-domains split across engineering and architecture are categorized as non-engineering.

[15] 47% of all North Carolina seniors took the SAT exam at least once in the second half of high school.

[16] 56% of the College Board sample of SAT test-takers are female, compared to 57% of UNC enrollees.

[17] There is evidence of sorting across the UNC system 16-campus system, with only 52% of females enrolling in UNC engineering campuses compared to 57% among all campuses. The main engineering campuses include North Carolina State University (NCSU), East Carolina University, North Carolina A&T State University, UNC - Charlotte, and Western Carolina University. Other campuses such as Appalachian State University and Elizabeth City State University offer limited engineering technology programs. NC State University is considered the state's premier engineering degree-granting institution. Its STEM-focused curriculum contrasts with the state flagship, UNC Chapel Hill. These subject differences in degree offerings arise in part from the evolution of land-grant and flagship universities common in many states.

North Carolina high school students and the national sample of college freshmen were asked about their first choice or probable major. Despite these differences, evidence exists to support an actual broadening of the gender disparity during high school. Using the same sample and question wording, the first HSLS follow-up in 2012 showed a gap of 9.8 percentage points among students expected to be in 11th grade.

Table 2 underscores the magnitude of differential entry during high school and the early postsecondary years. Another important factor to consider is the role of exit. Females may switch out at higher rates due to any combination of academic, social, and environmental factors. The longitudinal nature of the UNC sample provides a suitable context for examining attrition from engineering. I follow a sample of full-time, first-time freshmen enrollees in 2003-2010 whose initial declared major was engineering. Table 3 regresses an indicator for dropping out of engineering on gender and other covariates. The first column shows a raw attrition gap among females that is statistically insignificant from male peers. Conditioning on cohort fixed effects and academic performance as measured by SAT scores and high school GPA attenuates the still insignificant gender disparity in persistence. Taken together, the evidence suggests that higher attrition among aspiring women is not a first-order explanation for postsecondary gender disparities in engineering. Women have already sorted away from the field at the point these datasets begin documenting major intentions.[18]

⟨ Table 3 about here ⟩

5 Empirical approach

Stark gender disparities prompt the question of what affects students' decision to pursue engineering. In order to characterize inputs into the major choice decision that draw on common hypotheses cited in the literature, we can express the utility attached to engineering as:

$$U = f(X, A, S, \Lambda) \tag{1}$$

X is a set of sociodemographic characteristics such as gender, race, and parental occupations. A captures academic preparation and achievement reflecting individuals' ability to perform well in engineering-related coursework. This variable also encompasses the

[18]The present timespan precludes studying attrition in the labor market. As such, subsequent attrition from the engineering workforce due to professional, social, and family reasons may play a significant role but are outside the scope of this study.

effort exerted to lay the academic foundations for entering this major, as well as exposure to key math and science courses in secondary education or earlier that can shape students' engineering interests. S is a set of beliefs about ability shaping students' expectations for success. S is subject to interactions with individual attributes such as gender and ability, to the extent that parental and classroom influences shape ability beliefs differentially along individual characteristics and academic achievement measures. Finally, Λ describes preferences and work-related tastes such as pecuniary vs. non-pecuniary motivations and prosocial orientation. This furthermore reflects the cumulative effect of exposure to social context and its interactions with gender. The baseline model I estimate conditions on only a parsimonious set of individual characteristics X to characterize the raw gender gap in engineering:

$$Y_i = \alpha^{base} + \beta^{base} Fem_i + \phi^{base} X_i + \epsilon_i^{base} \tag{2}$$

Y_i indicates whether individual i expresses an engineering orientation or is observed to major in the subject. X is a set of individual characteristics such as race. The parameter of interest, β, expresses the average gender difference in engineering participation. In order to determine the contribution of ability beliefs, prosocial preferences, and other contextual factors to the gap, I augment the baseline model with several factors in the full specification:

$$Y_{ij} = \alpha^{full} + \beta^{full} Fem_i + \phi^{full} X_i + \gamma_1^{full} A_{ij} + \gamma_2^{full} S_{ij} + \gamma_3^{full} \Lambda_{ij} + \eta_j + \epsilon_{ij}^{full} \tag{3}$$

The subscript j denotes high school. A is a vector of academic preparation variables including SAT math, SAT verbal, high school GPA, and high school credits earned in reading, math, science, and computer programming. I assume that the effect of ability beliefs S is estimable via survey questions that elicit self-assessments of mathematical, academic, and writing abilities. Λ represents a vector of survey responses that capture pecuniary, other-regarding, work-based, and family preferences.[19] There are likely other inputs into the major decision-making process not captured by existing covariates. For instance, students with unobserved preferences for engineering coursework may select into different high schools offering contrasting academic environments that shape major orientation. I include high school fixed effects η_j to address selection on school-level

[19] An individual with higher math ability beliefs, ceteris paribus, will invest more in STEM-related human capital, which can further shape engineering-related ability beliefs and preferences for engineering coursework. Due to data constraints, this specification abstracts away from the dynamic relationships between human capital investment, ability beliefs, and preference formation.

unobservables and omitted variables.[20]

A hurdle in estimating the contribution of each covariate to β is that results are non-robust to the sequence in which they are added to the base regression. I rely on the decomposition technique of Gelbach (2016) to render the factor contributions order-invariant.[21] The decomposition relies on the sample omitted variable bias formula to explain the sensitivity underlying the relationship between β and included covariates. The portion of the gender gap explained by new explanatory variables is $\hat{\delta}_{Fem} = \hat{\beta}^{base} - \hat{\beta}^{full}$. This total difference is separable into k additional covariate groups:

$$\hat{\delta}_{Fem} = \sum_k \hat{\delta}_{k,Fem} = \sum_k \hat{\Gamma}_{k,Fem} \hat{\gamma}_k^{full} \tag{4}$$

This setup makes clear that $\hat{\delta}_{k,Fem}$, the contribution of the k-th covariate (group), is the product of two channels of influence. The first is the male-female difference in this factor after partialling out all other explanatory elements in the base regression. $\hat{\Gamma}_{k,Fem}$ is the coefficient on Fem from an auxiliary regression of the k-th covariate on all explanatory variables in the base model. The amount explained by SAT math scores, for instance, depends on the raw gender difference in this attribute after conditioning on the basic set of individual characteristics. The second channel $\hat{\gamma}_k^{full}$ reflects how correlated the k-th covariate is to the outcome under the full model. A sufficiently small coefficient associated with SAT math suggests that it will not be a meaningful contributor to the gender gap.

6 Results

Findings begin with the high school engineering gender gap and focus on the roles of students' academic achievement, curricular exposure, and extracurricular involvement. Since major intentions are given in the second half of high school, questions remain on whether these major preferences and gaps persist into postsecondary schooling. I turn to linked UNC data to examine the stability of major preferences during the transition to college. Data from the national CIRP Freshmen Survey and a sample of twins from North Carolina expands the set of potential explanatory factors to include ability beliefs, prosocial values, and family background and structure.

[20] I estimate the model using high school fixed effects in the College Board and UNC samples only because the CIRP Freshmen Survey does not elicit high school IDs.

[21] The approach of Gelbach (2016) generalizes the Oxaca-Blinder technique while ensuring path independency.

6.1 High school major orientation and academic achievement

Table 4 augments the unconditional linear probability model from Table 2 with standardized test scores, cumulative GPA during the first half of high school, earned credits in reading, math, physical science, and computer programming, and participation in high school extracurriculars. Controlling for only SAT performance reduces the gender gap from 14.7 to 13.3 percentage points, since students are positively selected on math scores to enter engineering and negatively selected on verbal performance. The addition of GPA indicators widens the gap to 13.6 percentage points to account for increased male engineering participation when elevating their grades to the same level as female peers. Meanwhile, controlling for earned credits in grades 9 and 10 leads to minimal change. Another source of potential differentiation along gender lines is the choice of high school extracurriculars. Accounting for extracurricular participation in computer activities, journalism, government, music, theater, dance, and ROTC attenuates the gap to 14.2 percentage points. Female participation is noticeably higher in the theater and arts, although aspiring female engineers are similarly distinguished as their male counterparts by their reduced involvement in these activities relative to same-gender peers. Participation in computer-related activities is strongly associated with higher engineering take-up rates among both genders.

⟨ Table 4 about here ⟩

Columns 7 and 8 in Table 4 display decomposition results for four contributing factors: credits earned, high school choice, SAT scores and GPA, and extracurricular activities. The full model explains 12.2% of the aggregate 14.7 percentage point disparity. Earned credits across multiple subjects in the first half of high school have little tangible effect on the gender divergence in engineering orientation, explaining 0.6% of the gap. The role of high school choice is even smaller at 0.3%. SAT scores and cumulative GPAs explain a further 7.5%. This magnitude is notably less than earlier studies. For example, nearly one-third of the engineering gender gap among college graduates in the 1989 entering cohort are explained via differences in SAT performance (Turner and Bowen, 1999).[22] This is consistent with the literature showing convergence in standardized test scores over recent decades. Finally, extracurricular participation explain 3.8%, with the largest contributor as participation in computer-related activities.[23]

[22]SAT scores may play a greater role in this study due to the composition of high-ability students in the College and Beyond Database. They originate from 12 institutions: Stanford, Yale, Princeton, Kenyon, Oberlin, Swarthmore, Hamilton, Williams, Wesleyan, Bryn Mawr, Smith, and Wellesley. On the other hand, the use of categorical variables for SAT scores in place of 10-point indicators may underestimate the contribution of SAT scores to the gender gap.

[23]15% of aspiring female engineers distinguish themselves early by engaging in computer-related ex-

6.2 Major orientation during the transition to college

Having established that high school major orientation already diverges along gender lines, I turn to the transition to postsecondary and explore two sets of questions. One is how these differences translate to actual college major choice, and the other examines the extent to which differential college application and enrollment behaviors among men and women may explain changes in major orientation during the transition. Linked North Carolina data permits tracking students as they make the transition from high school to one of UNC's 16 campuses. I begin with a sample of high school students that chose a STEM field as their preferred major and eventually enrolled in UNC during 2010. After attaining 30 credit hours at UNC, 61% of these STEM-oriented students persisted in this track by declaring a STEM major (Table 5). Men are more likely to choose engineering while women prefer other STEM fields such as biological sciences. Of students who switch out of STEM, the largest share chose a major in business, legal studies, or social sciences. Students inclined towards engineering in high school exhibit a similar pattern of persistence within STEM. Nearly one-third exit STEM for the social sciences, business, humanities, and other majors.

⟨ Table 5 about here ⟩

Among those aspiring to an engineering degree, 43% of all females end up declaring an engineering major compared to 50% of men. This gender gap is statistically significant when incorporating data from the 2009 cohort, suggesting that female high school students are less likely to convert engineering orientation into actual major choice.[24] This prompts further inquiries into what takes place during the transition to college. I investigate whether differences are explained by varying levels of certainty in one's initial major orientation, applications to UNC campuses with established engineering undergraduate programs, and enrollment in these campuses. Covariates such as campus selection are endogenous as they are likely guided by the same underlying factors that propel students into engineering majors. I simply use these covariates to describe how behavior during key decision points may diverge along gender lines.

Table 6 follows a sample of students from high school through UNC enrollment and subsequent major choice. Using engineering major declaration at the time of 30 earned credit hours as an outcome, the first specification replicates the unadjusted engineering

tracurriculars during grades 9 and 10, compared to 9% of non-engineers. Analogous statistics among males are 16% and 12%.

[24]In analyses not shown, 44% of women in the 2009-2010 enrollment cohorts continued onto an engineering major, or 7 percentage points fewer than male peers.

gender gap of 11.5 percentage points in Table 2. Controlling for high school engineering orientation decreases the gap to 4.0 percentage points, suggesting that the majority of the gender divergence in major choice is already determined by the second half of high school. Decomposition results quantify the share at 59.4%. In contrast, the gap remains largely stable when the certainty of initial major preferences or UNC application behaviors are taken into account. 0.8% is attributable to differences in the certainty of high school major choice. While male students who declare an engineering major exhibit similar levels of certainty as those who choose an alternate field of study, aspiring female engineers are less sure of their decision (Table A2). Finally, the choice of UNC campus contributes 8.9% to the overall gap. Despite applying to comparable sets of UNC institutions, female students are less likely to enroll in UNC campuses offering an engineering degree. Enrollment choices are not likely driven by lower acceptance rates at engineering-focused institutions, as engineering-oriented females have superior high school GPA and comparable SAT math scores. Other factors are prompting engineering-oriented female students to opt out of this track during the transition to college.

⟨ Table 6 about here ⟩

6.3 Ability beliefs, prosocial values and other preferences

A deeper understanding of these gender disparities is possible by investigating explanatory accounts not typically captured by administrative datasets. I turn to the CIRP Freshmen Survey to evaluate the importance of five factors: beliefs in academic abilities, pecuniary goals, other-regarding values, professional contributions in the arts and sciences, and family considerations. Table 7 decomposes the contributions of these accounts. The full model reduces the gender gap by 3.8 percentage points, or 27%, from the unconditional gap of 14.0 percentage points. Cross-gender differences in SAT scores and high school GPA contribute 4.8% to the overall variation, compared to 7.5% in the high school sample. Potential explanations for this disparity range from differences in student composition to classical measurement error that can be introduced via self-reported academic achievement scores. Neither pecuniary goals nor family considerations had a sizable impact. While men were more likely to elevate the importance of financial gain, pecuniary goals accounted for only 0.5% of the gender disparity. The small mean difference in men and womens survey responses led to family considerations contributing a negligible amount to the gender disparity.

⟨ Table 7 about here ⟩

Beliefs in academic, mathematical, and writing ability explain another 7.5%. The majority of this effect is driven by lower math ability beliefs among female students, conditional on academic performance. Since students who are confident along this dimension are more likely to select into engineering, equalizing females' math ability beliefs with males' bridges the gender gap by an amount that rivals the contribution of standardized test scores and GPA. Next I turn to two accounts focused on preferences: prosocial values and professional goals. Women are over-represented among those those who place greater importance on helping others in difficulty and influencing social values, with prosocial values explaining 7.9% of the gap.[25] Cross-gender differences in professional goals such as making a theoretical contribution to science and becoming accomplished in the arts explain an additional 6.5%. The contribution of academic ability beliefs relative to individual preferences is notable given the mixed evidence in the literature. Among the few gender gap studies that jointly focus on these attributes, Zafar (2013) found that the majority of the gap is explained by gender differences in preferences and expected enjoyment of studying in different fields. The relatively small contribution assigned to self-confidence stands in contrast to the prominent role occupied by ability beliefs in other studies (Valian, 1998; Antecol and Cobb-Clark, 2013; Leslie et al., 2015).[26] These results establish that academic ability beliefs matter alongside individual preferences for STEM participation.

Results from the full sample of college freshmen may disguise heterogeneous responses across individual attributes. I re-examine these patterns by ethnicity and math ability in recognition of potential interactions between these individual attributes and explanatory factors. For example, family and cultural backgrounds can differentially shape how females perceive their academic mastery, leading to different contributions of academic ability beliefs. Figure 1 divides the sample by race/ethnicity into white, African American, Hispanic, and Asian. The aggregate explanatory power of academic ability beliefs, prosocial preferences, professional goals, and SAT scores and GPA is the highest for white students at 28%. Hispanic and Asian students lag slightly behind, while the model accounts for only 17% of the African American gender disparity. Professional goals and academic performance play a less consequential role among black students because differences in academic performance and theoretical interests among black men and women are smaller compared to other ethnic groups. The resulting African American gender gap

[25] 27% of female respondents believe that helping those in difficulty is essential, compared to 17% of males.

[26] In Zafar (2013), the cumulative contribution of academic ability beliefs, reconciling work and family, and beliefs about future earnings is less than 5% of the aggregate engineering gap and statistically insignificant, compared to 27% explained by beliefs about coursework enjoyment and 60% by other preferences.

in engineering is smaller than that of white students, and this is driven by higher levels of engineering interest among females rather than lagging participation among men.[27]

⟨ Figure 1 about here ⟩

Segmenting the sample by SAT math scores shows that the share of explained variation increases in SAT math performance (Figure 2). Among low scorers, beliefs about mathematical ability account for 4-5% of cross-gender differences in engineering intentions, compared to over 7% among the highest SAT math scorers. Moreover, math ability beliefs comprise the majority of explanatory power contributed by all academic ability beliefs. This suggests that math anxiety and confidence is a particularly salient feature of the decision to enter engineering. Similar to ability beliefs, the contribution of professional goals is steadily increasing in SAT math scores. 3% of the gap among low math achievers are explained by gender differences in professional goals, compared to over 10% among the highest achievers. Among higher-scoring individuals, gender differences in the importance of making a scientific or artistic contribution widen and the correlations between engineering intentions and goals such as making theoretical contributions in science strengthen.

⟨ Figure 2 about here ⟩

6.4 Family background and structure

Yet another source of variation is parental influence and family context. Parental expectations for academic achievement and occupational choices are shown to differ by the child's gender.[28] These expectations may be informed by parents' professional experiences. Namely, gender-stereotypical career orientation may be muted in families where mothers work in math-related occupations and serve as professional role models for daughters. To test this possibility, I condition the models on mothers' and fathers' occupational categories in the CIRP Freshmen Survey. Table 7 shows that accounting for parental occupations increases the engineering gender gap by 0.2%, suggesting that professional experiences have limited scope for bridging gender differences in engineering orientation.

[27] 17.5% of African American males exhibit interest in engineering compared to 17.6% among white males. 5.2% of African American females express an engineering orientation at this stage compared to 3.5% among white females.

[28] Research shows, for instance, that parents on average have lower academic expectations for daughters (Fryer and Levitt, 2010).

Parental occupation belongs to a broader set of mechanisms that can shape gender-based norms with consequences for STEM orientation. Family background, teachers and peers can also establish and reinforce gender-based expectations of success in math- and science-oriented subjects, which in turn can lead boys and girls to diverge in their human capital investments. While it is difficult to isolate each individual influence, one means of gauging their cumulative effect is to compare the academic trajectories of otherwise similar children from different family compositions. I use a sample of same- and opposite-sex twins in North Carolina, under the assumption that sex composition is as good as random across families with twins. This assumption is challenged by a lack of information on twin zygosity in North Carolina data. While much empirical evidence supports the conjecture that fraternal (dizygotic) twin pairs are as likely to be same-sex as opposite-sex, I cannot rule out that genetic or environmental factors affect the occurrence of identical (monozygotic) twins included under the sample of same-sex twins.[29] The inability to restrict the sample to only dizygotic twins is taken into account when interpreting results. Under the assumption of random sex assignment among twin pairs, opposite-sex twins are distinguished by the salience of gender norms and any divergent engineering orientation on their part may be explained by differential gender role-based socialization during childhood and adolescence. When monozygotic twins are included in the sample, interpretation needs to consider the relationship between future STEM orientation and unobserved genetic or environmental factors shaping the family context for identical twins.

Table A4 shows the twins' engineering orientation at the end of high school while chronicling their academic achievement, attrition, SES, and computer use during elementary and middle school. The top panel juxtaposes engineering orientation and mean test scores of students from opposite sex twins with that of same-sex twins, conditional on taking the SATs. The bottom panel uses the full sample of 9569 twin pairs. The shares of college-aspiring females from both types of family structures aiming to major in engineering are statistically indistinguishable at 2%. On the other hand, males from opposite-sex twin pairs are significantly more likely to indicate an interest in engineering. 19% named engineering as their preferred major compared to 15% of same-sex male twins. Expressed in regression form, the unadjusted gap in Table 8 is 3.8 percentage points.[30]

[29]The Weinberg's differential rule establishes the independence of sexes in dizygotic twins. Multiple national registry datasets confirm the rule's robustness under different empirical contexts (Fellman and Eriksson, 2006).

[30]A specification of the following form is run separately for males and females:

$$Y_{ih} = \gamma + \delta\, OppSex_h + \rho\, X_{ih} + \epsilon_{ih} \tag{5}$$

⟨ Table 8 about here ⟩

The first candidate explanation for these patterns is academic preparation. Males in opposite-sex pairs may prefer engineering because they are better qualified in math than male-male pairs due to differential selection into the SAT sample or skill investment over time. In fact, the reverse is true - males in opposite-sex twin pairs are more interested in engineering despite having SAT math scores that are on average 10 points lower.[31] As a result, conditioning on SAT scores increases the advantage of opposite-sex male twins in engineering orientation from 3.8 to 4.2 percentage points (Table 8). Another explanation for these findings is that males in opposite-sex pairs are acting on their math advantage *relative* to their female twins. In an average pairing, males score almost 25 points higher in SAT math than females. I test the relative advantage hypothesis by restricting the control group to include only the twin in male-male pairs with higher math performance. In results not shown, males from opposite-sex pairs are still 2.6 percentage points more likely to choose engineering. Relative advantage cannot account for these differences.

One notable detail from Table A4 is that same-sex and opposite-sex pairs in the full sample consistently display similar levels of academic achievement, attrition rates, and socioeconomic status during middle school, with one exception. Males in opposite-sex pairings use computers more frequently than male-male twin pairings. Assuming these twins have similar access to home computers, the evidence is consistent with males in opposite-sex pairs investing more heavily in computer skills, which were previously shown to predict future engineering orientation. The exact mechanisms behind gender-based skill specialization can range from parental inputs to relationship dynamics within twin pairings, although they are difficult to determine in the absence of additional data on home environments and twin interactions.

These findings echo those from a study relating sibling sex composition to major choice. Conditional on attending a college-preparatory high school, males with at least one sister are more likely to choose a male-stereotypical major relative to males with only brothers (Anelli and Peri, 2015). This study's choice of twins stems from the need to address endogeneity concerns surrounding sibling sex composition. Twins are preferred over siblings because they are less affected by sex considerations in fertility decisions that are correlated with differential parental expectations and investments. Even then, the

Engineering orientation for individual i in household h depends on family structure, in this case whether the individual is part of an opposite-sex twin pair, and a vector of individual attributes such as standardized test scores (X_{ih}).

[31]The SAT math deficit among opposite-sex males is at least partially due to greater selection on ability into SAT test-taking among same-sex males. The latter group is 3 percentage points less likely to take the test, and those who do have significantly higher scores.

presence of identical twins in the sample suggests caution. Since identical twins share more genetic endowment, their skill investments and career orientation relative to the other may differ from same-sex fraternal twins. If these hard-to-observe differences in family context manifest in lower STEM-orientation, my results can overstate the contribution of gender role-based socialization and expectations.

7 Conclusion

Female under-representation in STEM fields such as engineering is a long-standing phenomenon. With the ascendance of women in college enrollment and completion, researchers are focusing on differential take-up and postsecondary attrition as key explanations for the STEM gender gap (Preston, 2004; Hunt, 2016). North Carolina longitudinal data used in this paper shows that selective attrition during higher education cannot explain the gap. Rather, pre-college differences and the transition to postsecondary merit additional scrutiny. Administrative and survey datasets document a sizable gender gap by 9th grade that persists throughout high school.[32] Efforts that promote STEM interest and increase exposure to engineering-related skills during secondary and elementary school require a better understanding of factors underlying engineering orientation. I decompose the gender gap into several explanatory accounts, including differences in academic performance, beliefs in ability, other-regarding values, and professional goals in the arts and sciences. SAT scores and high school GPA explain between 5 to 7% of the gap across three different datasets. High school sorting plays a negligible role, although high school curriculum choice, credits earned, and participation of extracurricular activities do matter for future academic tracks.

Beliefs in academic ability explain 8% of gender disparities in engineering, conditional on objective measures of academic ability. The majority of the result derives from differential beliefs in mathematical ability. Expectations about academic environment may be moderating this effect, as young women often hold themselves to higher standards in male-dominated fields like mathematics and engineering (Hill et al., 2010). The belief that they must be exceptionally good to succeed can reinforce the confidence gap and further exacerbate female under-representation (Correll, 2004). These confidence deficits have diverse social origins ranging from teachers' stereotypical biases to parental evaluations of competency (Herbert and Stipek, 2005; Gunderson et al., 2011). One means of

[32] Recent work on female under-representation in the academic sciences, for instance, corroborates the need to shift attention to pre-college major orientation over gender discrimination in the workforce (Ceci et al., 2014).

closing the engineering gap is to bridge gender differences in ability beliefs that are not justified by actual performance.

Remaining factors capture dimensions of individual preferences. Women are more likely to assign greater importance to values that correlate with lower engineering participation, including helping others in difficulty, influencing social values, and participating in community service work. These prosocial values collectively explain 8% of the gap. Professional goals in the arts and science explain a further 7%, with men disproportionately aiming to make a theoretical contribution to science. These associations suggest paths for future inquiry to better distinguish between underlying causes. Lower female participation may be driven by factors correlated with prosociality, such as appetite for risk and competition, preferences for job attributes such as the amount of collaborative teamwork, flexible work arrangements, mentorship opportunities, or preferences for applied vs. theoretical job tasks. Identifying the minimum set of sufficient conditions for behavioral change entails a more controlled setting that permits inference of causal relationships between these preferences and STEM participation.

Even when all observable influences are tallied, well over two-thirds of the gender participation gap in engineering remain. It is worth dwelling on the content of this unexplained residual. Candidate explanations span several categories, including family circumstances and parental inputs, preferences for the college experience, professional preferences, and labor market expectations. The twins-based analysis shows that opposite-sex pairs are more likely to pursue gender-stereotypical majors and suggests a role for parental investments and expectations. Although there exist challenges to causal interpretation in this context and limited information on the exact channels of influence, the prevalence of significant gender gap underscores the importance of studying gender-based socialization and expectations. Another component of the residual may be unmeasured preferences for college coursework and experiences that differ along gender lines. Gender differences in beliefs about coursework enjoyment explained over one-quarter of the engineering gender gap in a separate study (Zafar, 2013). Some of this divergence may be due to differential math ability beliefs, since students likely enjoy coursework more if they believe they would do well in the class. However, evidence suggests that other attributes, such as the receptiveness of the field to females, can shape beliefs about enjoyment. This relates to yet another possible point of divergence among men and women: professional preferences and labor market expectations. Decomposition results show that professional considerations feature prominently among high-ability students, who are disproportionately more likely to aspire to STEM careers. If otherwise well-qualified female students are more pessimistic about their opportunities for mentorship and advancement, this may explain part of the residual. Taken together, distinguishing between these conjectures require detailed longitudinal data on expectations, beliefs, inputs, and behaviors. The dynamic nature of human capital investment and preference formation suggests that laying the data groundwork can pay dividends for designing timely and tailored interventions.

References

Agostinelli, F. and Wiswall, M. (2016). Estimating the Technology of Children's Skill Formation. *National Bureau of Economic Research working paper #22442*.

Altonji, J. G., Blom, E., and Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, 4(1):185–223.

Anelli, M. and Peri, G. (2015). Gender of Siblings and Choice of College Major. *CESifo Economic Studies*, 61(1):53–71.

Antecol, H. and Cobb-Clark, D. A. (2013). Do psychosocial traits help explain gender segregation in young people's occupations? *Labour Economics*, 21:59–73.

Benbow, C. P. and Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: more facts. *Science*, 222(4627):1029–1031.

Bettinger, E. P. and Long, B. T. (2005). Do faculty serve as role models? the impact of instructor gender on female students. *The American Economic Review*, 95(2):152–157.

Beyer, S. and Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, 23(2):157–172.

Billger, S. M. (2002). Admitting men into a women's college: A natural experiment. *Applied Economics Letters*, 9(7):479–483.

Blau, F. D. and Kahn, L. M. (2000). Gender Differences in Pay. *Journal of Economic Perspectives*, 14(4):75–99.

Brown, C. and Corcoran, M. (1997). Sex-based differences in school content and the male-female wage gap. *Journal of Labor Economics*, 15(3):431–465.

Canes, B. J. and Rosen, H. S. (1995). Following in her footsteps? faculty gender composition and women's choices of college majors. *Industrial and Labor Relations Review*, 48(3):486–504.

Card, D. and Payne, A. A. (2017). High School Choices and the Gender Gap in STEM. Working Paper 23769, National Bureau of Economic Research.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125(3):1101–1144.

Ceci, S. J., Ginther, D. K., Kahn, S., and Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, 15(3):75–141.

Corbett, C. (2015). *Solving the equation: the variables for women's success in engineering and computing*. American Association of University Women, Washington, DC.

Correll, S. J. (2004). Constraints into Preferences: Gender, Status, and Emerging Career Aspirations. *American Sociological Review*, 69(1):93–113.

Dar-Nimrod, I. and Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science*, 314(5798):435–435.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, 42(3):528–554.

DiPrete, T. A. and Buchmann, C. (2013). *The Rise of Women: The Growing Gender Gap in Education and What it Means for American Schools*. Russell Sage Foundation.

Dweck, C. (2008). *Mindsets and math/science achievement*. New York: Carnegie Corporation of New York, Institute for Advanced Study, Commission on Mathematics and Science Education.

Dweck, C. S. (2000). *Self-theories: Their Role in Motivation, Personality, and Development*. Psychology Press.

Eccles, J. S., Jacobs, J. E., and Harold, R. D. (1990). Gender Role Stereotypes, Expectancy Effects, and Parents' Socialization of Gender Differences. *Journal of Social Issues*, 46(2):183–201.

Ellison, G. and Swanson, A. (2010). The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions. *Journal of Economic Perspectives*, 24(2):109–128.

Favara, M. (2012). The cost of acting "girly": Gender stereotypes and educational choices. Technical report, Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit.

Fellman, J. and Eriksson, A. W. (2006). Weinberg's differential rule reconsidered. *Human Biology*, 78(3):253–275.

Fryer, R. G. and Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2):210–240.

Gelbach, J. B. (2016). When Do Covariates Matter? And Which Ones, and How Much? *Journal of Labor Economics*, 34(2):509–543.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

Goldin, C., Katz, L., and Kuziemko, I. (2006). The homecoming of american college women: The reversal of the gender gap in college. *Journal of Economic Perspectives*, 20:133–156.

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review*, 29(6):911–922.

Gunderson, E. A., Ramirez, G., Levine, S. C., and Beilock, S. L. (2011). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, 66(3-4):153–166.

Herbert, J. and Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology*, 26(3):276–295.

Heyman, G. D. |. (2002). Gender and Achievement-Related Beliefs among Engineering Students. *Journal of Women and Minorities in Science and Engineering*, 8(1):41–52.

Hill, C., Corbett, C., and St. Rose, A. (2010). *Why so few?: women in science, technology, engineering, and mathematics*. American Association of University Women, Washington, D.C.

Hunt, J. (2016). Why do Women Leave Science and Engineering? *Industrial and Labor Relations Review Review*, 69(1):199–226.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, 321(5888):494–495.

Hyde, J. S. and Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, 106(22):8801–8807.

Kuncel, N. R., Credé, M., and Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, 75(1):63–82.

Legewie, J. and DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. *Sociology of Education*, 87(4):259–280.

Leslie, S.-J., Cimpian, A., Meyer, M., and Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219):262–265.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Niederle, M. and Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, 24(2):129–144.

Paglin, M. and Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics*, 8(1):123–144.

Preston, A. E. (2004). *Leaving Science*. Russell Sage Foundation.

Rothstein, D. S. (1995). Do female faculty influence female students' educational and labor market attainments? *Industrial and Labor Relations Review*, 48(3):515–530.

Sass, T. R. (2015). Understanding the STEM Pipeline. Working Paper 125. *National Center for Analysis of Longitudinal Data in Education Research (CALDER)*.

Sax, L. J. (1994). Mathematical self-concept: How college reinforces the gender gap. *Research in Higher Education*, 35(2):141–166.

Sax, L. J., Allison Kanny, M., Jacobs, J. A., Whang, H., Weintraub, D. S., and Hroch, A. (2016). Understanding the Changing Dynamics of the Gender Gap in Undergraduate Engineering Majors: 1971–2011. *Research in Higher Education*, 57(5):570–600.

Solnick, S. J. (1995). Changes in women's majors from entrance to graduation at women's and coeducational colleges. *Industrial and Labor Relations Review*, 48(3):505–514.

Turner, S. E. and Bowen, W. G. (1999). Choice of major: The changing (unchanging) gender gap. *Industrial and Labor Relations Review*, 52(2):289–313.

Valian, V. (1998). *Why So Slow? : The Advancement of Women*. MIT Press, Cambridge, Mass.

Vogt, C. M., Hocevar, D., and Hagedorn, L. S. (2007). A Social Cognitive Construct Validation: Determining Women's and Men's Success in Engineering Programs. *The Journal of Higher Education*, 78(3):337–364.

Xie, Y. and Shauman, K. A. (2003). *Women in science: career processes and outcomes*. Harvard University Press.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595.

Table 1: Share of recent college graduates in STEM fields

	STEM	Engineering	Computer Science	Biology	Math and Statistics	Physical Sciences	Engineer Tech	Agriculture	Science Tech
Female	14.8	2.5	0.9	7.0	0.9	2.5	0.2	0.9	0.1
Male	30.5	11.7	4.7	6.9	1.7	3.8	0.7	1.1	0.0

Notes: Sample comprises individuals from the 2008-2012 American Community Survey who graduated college and are between 22 - 23 years old.

Table 2: The engineering gender gap in high school and college

	High school		Postsecondary	
	HSLS: Gr. 9	College Board: Gr. 11 & 12	CIRP: freshmen	UNC: 30 credit hrs
	(1)	(2)	(3)	(4)
Female	-0.082^{***}	-0.147^{***}	-0.140^{***}	-0.115^{***}
	(0.003)	(0.002)	(0.000)	(0.006)
Black	-0.008	0.001	0.015^{***}	-0.037^{***}
	(0.005)	(0.002)	(0.001)	(0.006)
Hispanic	-0.013^{***}	0.009^{***}	0.021^{***}	-0.022^{*}
	(0.004)	(0.003)	(0.001)	(0.012)
Asian	0.028^{***}	0.026^{***}	0.044^{***}	0.009
	(0.009)	(0.003)	(0.001)	(0.017)
American Indian	-0.016	-0.011^{***}	-0.012^{***}	-0.043^{**}
	(0.022)	(0.004)	(0.005)	(0.020)
Other	-0.004	-0.012^{***}	0.006^{***}	0.012
	(0.006)	(0.003)	(0.001)	(0.015)
Observations	14668	266895	2042832	10548
R^2	0.040	0.070	0.072	0.045

Notes: HSLS:09 is a nationally representative sample of 9th graders in 2009. The outcome variable is constructed from weighted student responses to the question "what is the job or occupation that you expect or plan to have at age 30?" using the S1OCC30 variable. Students are categorized as being oriented towards engineering if they choose the "Architecture and Engineering" occupation category with the STEM sub-domain exclusively in Life and Physical Science, Engineering, Mathematics, and Information Technology Occupations. All students with STEM sub-domains split across engineering and architecture are categorized as non-engineering. Results are weight-adjusted. The College Board sample comprises students who took the SAT exam in 2009 - 2014 at least once as juniors or seniors. The outcome variable is constructed from a variable eliciting students' first choice major. The sample includes observations with non-missing SAT scores and those who can be linked to 9th or 10th grade transcript data. The base specification uses indicator variables for cohort year and academic level during the last SAT test administration. Standard errors are clustered at the high school level. The CIRP Freshmen Survey sample spans 1990-1999, 2001, 2004, 2006, 2008, and 2010 academic years. The outcome variable comes from a variable eliciting freshmen's probable field of study or major. The base specification uses year and college type indicators and student weights. The UNC sample comprises 2010 enrollees who declared a major by the time of attaining 30 credit hours at their home institution, with non-missing high school GPAs, SAT scores, and transcript data. Standard errors are clustered at the high school level. * p<0.1, ** p<0.05, *** p<0.01

Table 3: Attrition among engineering students

	Raw gap	Conditional gap	
Female	0.023	0.013	0.015
	(0.014)	(0.012)	(0.011)
Black		-0.049	-0.008
		(0.030)	(0.007)
Hispanic		-0.053^{**}	-0.048^{**}
		(0.018)	(0.015)
Asian		0.020	0.022
		(0.016)	(0.014)
American Indian		0.011	0.017
		(0.032)	(0.025)
Other		0.028	0.031
		(0.015)	(0.017)
SAT and high school GPA		Yes	Yes
Cohort FE		Yes	Yes
UNC campus FE			Yes
Observations	14889	14889	14889
R^2	0.000	0.036	0.040

Notes: Sample comprises students whose initial declared major is engineering. Dependent variable is an indicator for dropping out of engineering. The unadjusted model is then augmented with SAT math and verbal score indicators for each 10-point bin and for each high school GPA decile. * p<0.1, ** p<0.05, *** p<0.01

Table 4: High school gender gap

	OLS						Decomposition	
							Diff	% Explained
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.147^{***}	-0.146^{***}	-0.136^{***}	-0.146^{***}	-0.142^{***}	-0.129^{***}		
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)		
Earned credits in Grades 9 & 10							-0.001^{***}	-0.6%
Reading credits in Gr. 9				-0.004^{**}		-0.006^{***}		
				(0.002)		(0.002)		
Rath credits in Gr. 9				-0.002		0.003^{**}		
				(0.002)		(0.001)		
Physical science credits in Gr. 9				0.012^{***}		0.004		
				(0.003)		(0.003)		
Computer prog. credits in Gr. 9				0.032^{***}		0.014		
				(0.011)		(0.013)		
Reading credits in Gr. 10				-0.007^{***}		-0.007^{***}		
				(0.001)		(0.001)		
Math credits in Gr. 10				0.015^{***}		0.009^{***}		
				(0.002)		(0.001)		
Physical science credits in Gr. 10				0.021^{***}		0.002		
				(0.002)		(0.001)		
Computer prog. credits in Gr. 10				0.045^{***}		0.028^{***}		
				(0.010)		(0.011)		
High school fixed effects		Yes				Yes	-0.000^{*}	-0.3%
SAT and high school GPA			Yes			Yes	-0.011^{***}	-7.5%
Extracurriculars in Grades 9 and 10					Yes	Yes	-0.006^{***}	-3.8%
Observations	266895	266895	266895	266895	266895	266895		
R^2	0.070	0.077	0.091	0.072	0.072	0.101		
Total							-0.018^{***}	-12.2%

Notes: High school sample includes students who took the SAT exam in 2009 - 2014 at least once as juniors or seniors. The sample excludes observations with missing SAT scores and those who cannot be linked to College Board or transcript data. All OLS specifications include indicators for race, cohort, and the grade level of the latest SAT administration. Augmented models include earned credits in grades 9 and 10, high school fixed effects, and indicators for each 10-point SAT math and verbal score bin, deciles for cumulative GPA during grades 9 and 10, and extracurriculars spanning participation in computer activities, music/vocal, theater, junior ROTC, dance, government/political, and journalism/literary activities during both 9th and 10th grades. Robust standard errors are clustered at the school level. The last two columns decompose the 1.8 percentage point difference between the parsimonious model (Column 1) and full model (Column 6) into its constituent parts. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 5: College major choice among STEM-oriented high school students

College major choice	High school STEM orientation			Engineering orientation		
	All	Female	Male	All	Female	Male
Engineering	0.27	0.13	0.33***	0.49	0.43	0.50
Other STEM	0.34	0.43	0.30***	0.15	0.18	0.14
Art/Humanities	0.04	0.05	0.04	0.04	0.04	0.04
Health	0.02	0.05	0.01***	0.01	0.01	0.01
Education	0.03	0.05	0.02***	0.02	0.05	0.02**
Business/Legal/Social Sciences	0.16	0.16	0.16	0.13	0.07	0.14**
Other	0.14	0.13	0.14	0.16	0.21	0.15**
Observations	2233	692	1541	1050	149	901

Notes: The sample includes students who chose STEM as a first choice major in high school and enrolled in UNC in 2010. The sample excludes those who did not declare a major by the time they attained 30 credit hours at their home institution, and those with non-missing high school GPAs, SAT scores, or transcript data. The first 3 columns use the full sample, and the subsequent 3 columns condition on choosing engineering as the preferred major during high school. Stars denote statistically significant difference in means relative to students of the opposite gender. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 6: Transition to college engineering major

	OLS						Decomposition	
							Diff	% Explained
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.115^{***}	-0.040^{***}	-0.113^{***}	-0.112^{***}	-0.100^{***}	-0.035^{***}		
	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.005)		
High school engineering orientation		Yes				Yes	-0.068^{***}	-59.4%
Certainty of major preferences			Yes			Yes	-0.001^{**}	-0.8%
UNC application behavior				Yes		Yes	-0.000	-0.1%
UNC campus fixed effects					Yes	Yes	-0.010^{***}	-8.9%
Observations	10548	10548	10538	10548	10548	10538		
R^2	0.045	0.230	0.046	0.054	0.124	0.270		
Total							-0.079^{***}	-69.3%

Notes: The outcome is a dichotomous variable for declaring an engineering major after earning 30 credit hours in the UNC institution, conditional on not having dropped out. The sample includes 2010 UNC enrollees who declared a major by the time of attaining 30 credit hours at their home institution, with non-missing high school GPAs, SAT scores, and transcript data. The first OLS specification is the raw gender difference in actual college engineering major choice. Augmented models include indicators for picking engineering as the preferred major during high school, the certainty of high school major preferences (very certain, fairly certain, and not certain), an indicator for applying to at least one UNC campus that offers an engineering undergraduate program, and UNC campus fixed effects. Robust standard errors are clustered at the high school level. The last two columns decompose the 7.9 percentage point difference between the parsimonious model (Column 1) and full model (Column 6) into its constituent parts. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 7: Decomposition - CIRP Freshmen Survey

	Contribution	Share of gap
SAT and high school GPA	-0.007***	-4.8%
Self-confidence: academic ability	-0.011***	-7.5%
Self Rating: Academic ability		
Self Rating: Mathematical ability		
Self Rating: Writing ability		
Pecuniary goals	-0.001***	-0.5%
Goal: Being very well off financially		
Social and other-regarding values	-0.011***	-7.9%
Goal: Helping others who are in difficulty		
Goal: Influencing social values		
Future act: Participate in volunteer/community service		
Professional goals in the arts and sciences	-0.009***	-6.5%
Goal: Becoming accomplished in one of the perf. arts		
Goal: Creating artistic work (painting, sculpture, etc.)		
Goal: Making a theoretical contribution to science		
Family considerations	0.000	0.0%
Goal: Raising a family		
Parental occupations	0.000***	0.2%
Father's aggregate career category		
Mother's aggregate career category		
Number of college applications	-0.000***	-0.1%
Total	-0.038***	-27.1%

Notes: CIRP Freshmen Survey sample spans 1990-1999, 2001, 2004, 2006, 2008, and 2010 academic years. The inclusion of the full set of covariates reduces gender gap from 14.0% to 10.2%. Specifications include year and college type fixed effects, and use student weights. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 8: Engineering orientation among twin sample

	Females		Males	
Opposite-sex twin	-0.006	-0.006	0.039^{***}	0.042^{***}
	(0.005)	(0.005)	(0.013)	(0.013)
Race and cohort FE		Yes		Yes
SAT scores		Yes		Yes
Observations	4420	4420	3571	3571
R^2	0.000	0.064	0.003	0.059

Notes: Twins are flagged in the NCERDC on the basis of identifying information such as birth date, name, and home address. The full sample includes all twins appearing in grade 3 during 2000 - 2005, up to grade 8 in 2005 - 2010, that have non-missing SAT score information between 2009 - 2014. The female sample comprises both females in opposite-sex twins and same-sex female twins. The male sample comprises both males in opposite-sex twins and same-sex male twins. Models are augmented with indicators for SAT math and verbal scores for each 10-point bin, grade at the latest SAT test administration, and cohort fixed effects. * p<0.1, ** p<0.05, *** p<0.01

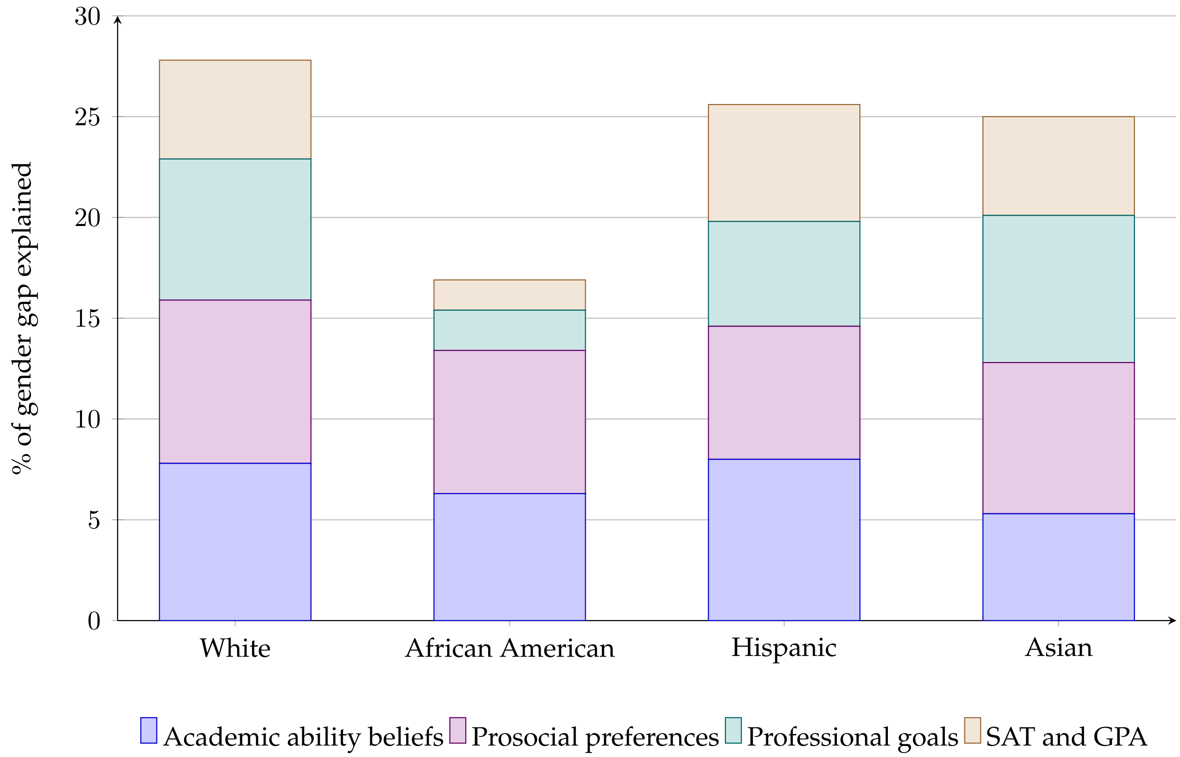

Figure 1: Decomposition by ethnicity

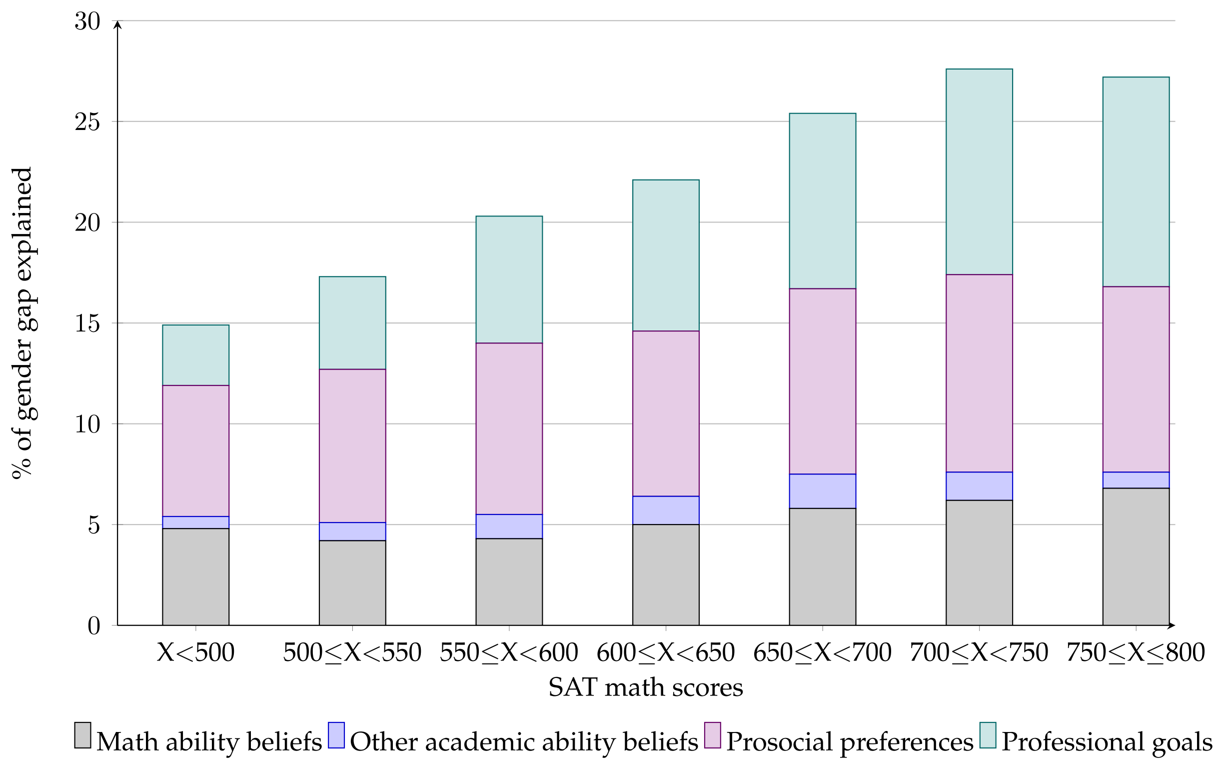

Figure 2: Decomposition by math ability

Appendix

A Summary statistics

Table A1: Summary statistics - high school sample

	Female		Male	
	Engineer	Not engineer	Engineer	Not engineer
SAT math	568.10	493.49***	553.99	517.08***
SAT verbal	533.42	487.90***	501.83	491.92***
Cumulative GPA in Grades 9 and 10	3.82	3.41***	3.44	3.24***
Earned reading credits in Grade 9	1.18	1.18	1.14	1.16***
Earned math credits in Grade 9	1.22	1.24**	1.23	1.24***
Earned physical science credits in Grade 9	0.08	0.07**	0.07	0.07**
Earned computer programming credits in Grade 9	0.01	0.00***	0.01	0.00***
Earned reading credits in Grade 10	1.11	1.15***	1.07	1.09***
Earned math credits in Grade 10	1.26	1.17***	1.20	1.16***
Earned physical science credits in Grade 10	0.53	0.34***	0.39	0.33***
Earned computer programming credits in Grade 10	0.01	0.00***	0.02	0.01***
Grade 9 computer activity	0.15	0.09***	0.16	0.12***
Grade 10 computer activity	0.15	0.09***	0.16	0.12***
Grade 9 music/vocal activity	0.08	0.11***	0.03	0.04***
Grade 10 music vocal activity	0.08	0.11***	0.03	0.04***
Grade 9 theater activity	0.08	0.09**	0.03	0.04***
Grade 10 theater activity	0.08	0.10***	0.03	0.05***
Grade 9 junior ROTC	0.04	0.02***	0.05	0.04***
Grade 10 junior ROTC	0.04	0.02***	0.05	0.04***
Grade 9 dance activity	0.11	0.12	0.01	0.01***
Grade 10 dance activity	0.11	0.12**	0.01	0.01***
Grade 9 govt/political activity	0.08	0.05***	0.03	0.03***
Grade 10 govt/political activity	0.11	0.07***	0.04	0.05***
Grade 9 journalism/literary activity	0.03	0.03	0.01	0.01***
Grade 10 journalism/literary activity	0.05	0.06**	0.01	0.02***
Observations	2936	144344	20031	99584

Notes: high school sample comprises students who took the SAT exam in 2009 - 2014 at least once as juniors or seniors. Sample excludes observations with missing SAT scores and those who cannot be linked to 9th or 10th grade transcript data. Stars denote statistically significant difference in means relative to students of same gender group who chose engineering as preferred major. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A2: Summary statistics - UNC sample

	Female		Male	
	Engineer	Not engineer	Engineer	Not engineer
SAT math	597.77	524.27***	605.07	561.93***
SAT verbal	551.93	518.83***	539.42	529.58***
High school GPA	4.23	3.80***	3.99	3.71***
High school engineering orientation	0.27	0.01***	0.64	0.12***
Very certain in first choice major	0.20	0.36***	0.26	0.27
Applied to at least 1 of 5 main UNC engineering campuses	0.98	0.79***	0.94	0.81***
Enrolled in UNC engineering campus	1.00	0.49***	0.96	0.55***
Observations	233	5803	702	3810

Notes: Sample includes 2010 UNC enrollees who declared a major by the time of attaining 30 credit hours at their home institution, with non-missing high school GPAs, with non-missing high school GPAs, SAT scores, and transcript data. The 5 main UNC engineering campuses include North Carolina State University (NCSU), East Carolina University, North Carolina A&T State University, UNC - Charlotte, and Western Carolina University. A small share of students applied to and enrolled in campuses such as Appalachian State University and Elizabeth City State University, which offer limited engineering programs. Stars denote statistically significant difference in means relative to students of the opposite gender. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A3: Summary statistics - CIRP Survey

	Female		Male	
	Engineer	Not engineer	Engineer	Not engineer
SAT math	630.20	555.14***	631.80	581.88***
SAT verbal	579.35	546.65***	565.59	549.94***
Self Rating: Academic ability	4.28	3.91***	4.24	4.00***
Self Rating: Mathematical ability	4.15	3.35***	4.18	3.61***
Self Rating: Writing ability	3.47	3.56***	3.36	3.52***
Goal: Being very well off financially	3.02	2.96***	3.16	3.09***
Goal: Helping others who are in difficulty	2.79	2.97***	2.55	2.70***
Goal: Influencing social values	2.18	2.42***	2.03	2.29***
Future Act: Participate in volunteer/comm. service work	3.11	3.11***	2.61	2.67***
Goal: Becoming accomplished in one of the performing arts	1.46	1.60***	1.36	1.52***
Goal: Creating artistic work (painting, sculpture, etc.)	1.42	1.57***	1.37	1.51***
Goal: Making a theoretical contribution to science	2.24	1.67***	2.18	1.77***
Goal: Raising a family	2.94	3.07***	3.04	3.07***
Observations	41624	1042214	168047	790947

Notes: Freshmen Survey sample spans 1990-1999, 2001, 2004, 2006, 2008, and 2010 academic years. Means reported using sample weights. Self-rating variables are reported on a scale of 1-5: 1) lowest 10%, 2) below average, 3) average, 4) above average, and 5) highest 10%. Goals and Future Acts are reported on a 1-4 scale. The corresponding categories for goals are: 1) not important, 2) somewhat important, 3) very important, and 4) essential. The scale for future acts inquire about the probability of undertaking a future action: 1) no chance, 2) very little chance, 3) some chance, and 4) very good chance. Stars denote statistically significant difference in means relative to students of same gender group who chose engineering as preferred major. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A4: Summary statistics - twin sample

	Female				Male			
	Mean		N		Mean		N	
	Same-sex	Opp-sex	Same-sex	Opp-sex	Same-sex	Opp-sex	Same-sex	Opp-sex
Conditional on taking SATs								
Engineering orientation	0.02	0.02	2827	1593	0.15	0.19***	2256	1315
SAT math	491.58	491.80	2827	1593	525.14	515.66**	2256	1315
SAT verbal	484.76	485.92	2827	1593	493.51	485.68**	2256	1315
Full sample								
Engineering orientation	0.01	0.01	6174	3386	0.05	0.07***	6194	3386
Took SATs	0.46	0.47	6174	3386	0.36	0.39**	6194	3386
Math EOG in Grade 8	0.15	0.12	5130	2838	0.11	0.09	4943	2746
Math EOG in Grade 7	0.11	0.09	5291	2903	0.10	0.08	5088	2820
Math EOG in Grade 6	0.09	0.09	5301	2946	0.07	0.06	5153	2890
Reading EOG in Grade 8	0.15	0.12	5130	2839	0.05	0.03	4911	2743
Reading EOG in Grade 7	0.17	0.16	5287	2902	0.02	0.01	5059	2811
Reading EOG in Grade 6	0.14	0.14	5291	2943	0.00	-0.01	5119	2885
Attrited in Gr. 7	0.05	0.04*	6174	3386	0.05	0.04	6194	3386
Attrited in Gr. 6	0.03	0.03	6174	3386	0.03	0.03	6194	3386
Attrited in Gr. 5	0.02	0.02	6174	3386	0.03	0.03	6194	3386
Free or reduced lunch in Grade 8	0.42	0.41	5134	2843	0.41	0.40	4994	2754
Free or reduced lunch in Grade 7	0.42	0.40	5080	2782	0.40	0.38	4947	2710
Free or reduced lunch in Grade 6	0.43	0.42	5178	2856	0.41	0.41	5087	2832
Use computer $\geq$ 1-2 times/wk in Gr 8	0.35	0.34	5112	2825	0.28	0.31***	4919	2731
Use computer $\geq$ 1-2 times/wk in Gr 7	0.27	0.29*	5267	2887	0.22	0.25***	5077	2801
Use computer $\geq$ 1-2 times/wk in Gr 6	0.23	0.24	5300	2940	0.19	0.21**	5154	2876

Notes: sample comprises all twin pairs appearing in grade 3 during 2000 - 2005, up to grade 8 in 2005 - 2010. Twins are flagged in NCERDC on the basis of identifying information such as birth date and home address. The full sample includes 3386 opposite-sex pairs, 3087 female same-sex pairs, and 3097 male same-sex pairs. The top panel restricts to individuals who took the SATs and have non-missing score information between 2009 - 2014. Stars denote statistically significant difference in means of students in a given gender group who belong to a same-sex versus opposite-sex pair. * p<0.1, ** p<0.05, *** p<0.01

B Sample selection

Table B1: High school senior and College Board samples

	Full sample	College Board sample
Female	0.51	0.55***
Black	0.27	0.25***
Hispanic	0.08	0.05***
Asian	0.03	0.03***
American Indian	0.01	0.01***
White	0.58	0.63***
Other	0.03	0.03***
Economically disadvantaged	0.38	0.27***
Observations	545534	255221

Notes: The full sample includes all seniors enrolled in a North Carolina public or charter high school in 2009-2014. The College Board sample includes all matched seniors in 2009-2014 who took the SATs in their junior or senior year with non-missing transcript data. Stars denote statistically significant difference in means from full high school sample. * p<0.1, ** p<0.05, *** p<0.01

Table B2: College Board and UNC samples

	College Board sample		UNC sample	
	Full	Applied to engi UNC campuses	Full	Enrolled in engi UNC campus
Female	0.56	0.55***	0.57***	0.52***
Black	0.25	0.23***	0.26**	0.24***
Hispanic	0.04	0.03***	0.03***	0.03***
Asian	0.03	0.04***	0.04***	0.04***
American Indian	0.01	0.01***	0.01	0.01***
White	0.64	0.66***	0.65**	0.67***
Other	0.02	0.02**	0.02***	0.02*
Economically disadvantaged	0.25	0.22***	0.20***	0.19***
Observations	41820	25482	19198	9744

Notes: The full College Board sample includes all seniors in 2010 who took the SATs in their junior or senior year with non-missing transcript data. The next sample includes all students who applied to one of five UNC campuses offering an undergraduate engineering degree: East Carolina University, North Carolina A&T State University, UNC - Charlotte, North Carolina State University, and Western Carolina University. The College Board NCSU sample only includes students who applied to North Carolina State University. The UNC sample includes all students matched to the full 2010 high school sample who enrolled in one of 16 UNC campuses. The last column shows the subset of these students who enrolled in an engineering campus. Stars denote statistically significant difference in means from full College Board sample. * p<0.1, ** p<0.05, *** p<0.01